IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Willy Fu
30/12/2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

This project focuses on predicting the successful landing of SpaceX Falcon 9's first stage, a crucial step towards making space exploration more cost-effective through reusable rockets. The project was carried out in several stages, combining data collection, analysis, visualization, and machine learning techniques.

<u>Methodologies</u>

- **Data collection:** Leveraged the SpaceX API to obtain detailed launch data, including payloads, launch sites, and outcomes. Supplemented with web scraping from Wikipedia to enhance the dataset.
- **Data wrangling:** Performed cleaning and integration of datasets to ensure consistency and usability.
- **Exploratory Data Analysis (EDA):** Conducted SQL queries to identify trends and patterns in launch outcomes. Created visualizations, such as scatter plots and bar charts, to understand relationships in the data.
- **Interactive visual analytics:** Generated Folium maps to analyze geographic features around launch sites. Developed Plotly Dash dashboards for interactive exploration of launch metrics.
- **Predictive modeling:** Built machine learning models to predict first-stage landing success. Evaluated models through metrics like accuracy and confusion matrices.

Executive Summary

Summary of all results

- SpaceX has demonstrated a steady improvement in success rates since 2013, achieving milestones in 2017 and 2019, showcasing advancements in reusable rocket technology.
- KSC LC-39A accounts for 41.7% of successful launches, highlighting its strategic importance.
- The Falcon 9 Block 5 booster consistently delivers high reliability for heavy payloads.
- Predictive models identified the Decision Tree as the most accurate (89%) for landing success.

Introduction

Project background and context: The SpaceX Falcon 9 rocket is a reusable spacecraft designed to reduce the cost of space exploration. A key challenge in achieving reusability is ensuring the successful landing of the rocket's first stage. To address this, SpaceX has conducted numerous launches with varying levels of success. By analyzing historical launch data, we aim to uncover patterns and factors that contribute to successful landings, providing SpaceX with actionable insights to refine their launch strategies.

Problems I wan to find answers to:

- What are the key factors that influence the successful landing of the Falcon 9 first stage?
- How do payload mass, orbit type, and other variables impact the likelihood of success?
- Can machine learning models reliably predict the outcomes of future launches?
- How can interactive visualizations aid in understanding and communicating launch data effectively?

Section 1

Methodology

Methodology

Executive Summary

1. **Data collection methodology:**

   - **API Calls:**
     - Utilized SpaceX's REST API to retrieve comprehensive data on launch details, including payload mass, orbit type, launch sites, and outcomes.
     - Extracted structured JSON data and transformed it into tabular formats for analysis.
   - **Web Scraping:**
     - Supplemented the dataset by scraping Wikipedia for additional launch metadata, such as mission objectives and payload descriptions.

2. **Data wrangling and processing:**

   - Addressed missing values by using domain knowledge to fill gaps where appropriate.
   - Standardized units and formats across datasets to ensure compatibility.
   - Merged datasets collected from different sources into a single cohesive database.

Methodology

3. **Exploratory Data Analysis (EDA):**
   - **Conducted SQL queries to identify critical patterns:**
     - Success rates by orbit type.
     - Average payload masses for specific booster versions.
   - **Used visualizations, such as scatter plots, bar charts, and time-series graphs, to examine relationships and trends.**
4. **Interactive visual analytics:**
   - **Folium maps:**
     - Created interactive maps displaying launch site locations and success rates.
     - Highlighted geographic proximities to infrastructure like highways and coastlines.
   - **Plotly Dash dashboards:**
     - Built dynamic dashboards to explore variables like payload vs. launch outcomes.
     - Integrated range sliders and drop-down menus for user interactivity.

Methodology

5. **Predictive analysis using classification models:**

   - **Model building:**
     - Trained multiple machine learning models (e.g., logistic regression, decision trees) to predict landing outcomes.
   - **Tuning and evaluation:**
     - Performed hyperparameter tuning to optimize model performance.
     - Evaluated models using metrics like accuracy, precision, recall, and F1-score.
   - **Model insights:**
     - Visualized feature importance to identify key variables impacting predictions.

Data Collection

How data sets were collected:

1. **API data retrieval:**
   - Leveraged SpaceX's REST API to collect structured data on Falcon 9 launches, including:
     - Launch dates, payload details, orbit types, and outcomes.
     - Booster information and success rates for landing attempts.
   - The data was retrieved in JSON format and transformed into tabular datasets for further analysis.
2. **Web scraping:**
   - Supplemented API data with web scraping from Wikipedia to obtain:
     - Historical mission descriptions, additional payload metadata, and detailed launch objectives.
     - Scraping was performed using Python libraries such as BeautifulSoup to parse HTML content.
3. **Data integration and consolidation:**
   - Combined data from both sources into a unified dataset, ensuring consistency and accuracy through data wrangling techniques.

Data Collection

Data collection process flowchart:

API data retrieval

- Initiate API calls to SpaceX's endpoint to fetch raw JSON data.
- Parse and validate the data to handle missing or incomplete fields.

Web scraping workflow

- Identify target Wikipedia pages for Falcon 9 launches.
- Extract tables and text-based content related to mission details.
- Clean and organize scraped data into a structured format.

Data integration and consolidation

- Perform data cleaning operations, such as deduplication and standardization.
- Merge datasets from API and scraping workflows into a single comprehensive database.

Data Collection – SpaceX API

- Falcon 9 launches data collection with SpaceX REST calls:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedC
4	1	2010-06-04	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0
5	2	2012-05-22	Falcon 9	525.000000	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0
6	3	2013-03-01	Falcon 9	677.000000	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0
7	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0
8	5	2013-12-03	Falcon 9	3170.000000	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-1/1_jupyter-labs-spacex-data-collection-api.ipynb

Initiate API call

Send GET requests to SpaceX API endpoint.

Validate response

Check HTTP status codes and JSON structure.

Extract relevant data

Parse key metrics from JSON.

Transform and store

Format data into tabular structure

Data Collection – Scraping

Web scraping process:

Identifying the data source

- The Wikipedia page *"List of Falcon 9 and Falcon Heavy launches"* was selected as the primary source of historical data on launches.

HTTP request and content retrieval

- Used Python's requests library to perform a GET request to the target page URL.
- Retrieved the HTML content of the page for further processing.

Parsing the HTML content

- Leveraged BeautifulSoup to parse the HTML content and locate the table containing launch records.

Data extraction and cleaning

- Iterated over the table rows to extract relevant information.
- Clean and normalize the extracted data, handling null values and ensuring uniform formats.

Data storage

- Organized the processed data into a Pandas DataFrame, enabling streamlined analysis and visualization.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-1/2_jupyter-labs-webscraping.ipynb

Data Wrangling

Data wrangling process:

Cleaning the data

- Identified and handled missing values.
- Removed duplicate entries to ensure data integrity.

Data standardization

- Unified formats for key variables such as dates, payload mass, and orbit types.
- Converted units where necessary to maintain consistency.

Integration of data sources

- Merged datasets collected from the SpaceX API and web scraping efforts.
- Resolved inconsistencies by cross-referencing overlapping data points.

Feature engineering

- Created new variables to enhance analysis, such as calculating success rates by launch site and deriving binary labels for landing outcomes.

Validation

- Performed sanity checks to ensure accuracy and reliability of the processed data.
- Used summary statistics and visual inspections to validate the dataset.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-1/3_labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

<u>Summary of charts and their purpose:</u>

1. **Relationship between Flight Number and Launch Site:**
   - **Chart type**: Scatter Plot
   - **Purpose**: To analyze how the sequence of launches correlates with specific launch sites and identify any patterns or trends in launch site usage over time.
2. **Relationship between Payload Mass and Launch Site:**
   - **Chart type**: Scatter Plot
   - **Purpose**: To understand how payload mass varies across different launch sites and assess if certain sites handle larger or smaller payloads consistently.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-2/5_edadataviz.ipynb

EDA with Data Visualization

3. **Success rate of each Orbit Type:**
   - **Chart Type:** Bar Chart
   - **Purpose:** To visualize the success rates for each orbit type and determine which orbits have the highest and lowest success probabilities.
4. **Relationship between Flight Number and Orbit Type:**
   - **Chart Type:** Scatter Plot
   - **Purpose:** To examine trends in orbit types over sequential launches and explore how orbit selection might influence launch success.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-2/5_edadataviz.ipynb

EDA with Data Visualization

5. **Relationship between Payload Mass and Orbit Type:**
   - **Chart Type:** Scatter Plot
   - **Purpose:** To explore the distribution of payload masses for different orbit types and investigate potential constraints or preferences related to payload size.
6. **Launch success yearly trend:**
   - **Chart Type:** Line Chart
   - **Purpose:** To track the annual success rate of launches and identify improvements or declines in performance over the years.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-2/5_edadataviz.ipynb

EDA with SQL

Summary of SQL queries performed:

- Retrieve all launch site names:

```
result = %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
```

- Filter launch sites starting with “CCA”:

```
result = %sql SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

- Calculate total payload carried by NASA boosters:

```
result = %sql SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass FROM
SPACEXTBL WHERE "Customer" = 'NASA (CRS)';
```

- Compute average payload mass for booster version F9 v1.1:

```
result = %sql SELECT AVG("Payload_Mass__kg_") AS Average_Payload_Mass FROM
SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1';
```


GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-2/4_jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

- Identify the date of the first successful ground landing:

```
result = %sql SELECT MIN("Date") AS First_Successful_Landing FROM SPACEXTBL
WHERE "Landing_Outcome" = 'Success (ground pad)';
```

- List boosters with successful drone ship landings and specific payload range:

```
result = %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" =
'Success (drone ship)' AND "Payload_Mass__kg_" > 4000 AND "Payload_Mass__kg_" < 6000;
```

- Count total number of successful and failed missions:

```
result = %sql SELECT "Mission_Outcome", COUNT(*) AS Total_Count FROM
SPACEXTBL GROUP BY "Mission_Outcome";
```

- Find boosters with maximum payload mass:

```
result = %sql SELECT "Booster_Version" FROM SPACEXTBL WHERE "Payload_Mass__kg_" = (SELECT
MAX("Payload_Mass__kg_") FROM SPACEXTBL);
```


GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-2/4_jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

- List failed drone ship landings in 2015:

```
result = %sql SELECT SUBSTR("Date", 6, 2) AS Month, "Booster_Version", "Launch_Site", "Landing_Outcome"
FROM SPACEXTBL WHERE "Landing_Outcome" = 'Failure (drone ship)' AND SUBSTR("Date", 1, 4) = '2015';
```

- Rank landing outcomes (2010-2017):

```
result = %sql SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE "Date"
BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY Outcome_Count DESC;
```

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-2/4_jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- **Markers:**
  - Added markers for each launch site to clearly identify their geographic locations and customized markers with labels displaying the launch site names for clarity.
  - Essential for pinpointing exact locations of launch sites on the map, aiding in visual orientation.
- **Circle markers:**
  - Used circle markers to represent the success rates of launches from each site. The size and color of the circles were scaled to reflect the success rates visually.
  - Added to give a visual representation of launch success rates, making it easier to compare performance across sites.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-3/6_lab_jupyter_launch_site_location.ipynb

Build an Interactive Map with Folium

- **Lines:**
  - Incorporated lines to illustrate proximities between launch sites and nearby infrastructure, such as highways, railways, and coastlines.
  - Used to analyze and display the geographic relationship between launch sites and surrounding infrastructure, which could impact logistical planning.
- **Pop-ups:**
  - Integrated pop-ups for each marker to provide additional details, such as the total number of launches and success statistics for the corresponding site.
  - Enhance interactivity by providing detailed information without cluttering the map, making it more user-friendly.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-3/6_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- **Launch success pie chart:**
  - Displays the proportion of successful and failed launches for all sites.
  - **Interactivity:** Filters can dynamically update the chart based on specific launch sites.
  - **Explanation:** Added to provide a quick and clear overview of the overall success rate of SpaceX launches, enabling users to assess mission reliability.
- **Payload mass vs. launch outcome scatter plot:**
  - Plots payload mass against launch outcomes for all launch sites.
  - Interactivity: Includes range sliders to filter payload mass ranges and dropdowns to select specific orbit types.
  - **Explanation:** Designed to investigate how payload mass affects launch outcomes and to identify optimal payload ranges for successful missions.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-3/6_bis_spacex_dash_app_v2.py

Build a Dashboard with Plotly Dash

- **Launch success rate by site bar chart:**
    - Highlights the success rate of launches for each launch site.
    - Interactivity: Responds to user inputs for site selection to provide detailed statistics.
    - **Explanation:** Included to compare performance across different launch sites, helping stakeholders focus on the most reliable locations.
- **Launch trends line chart:**
    - Tracks yearly trends in launch success rates.
    - Interactivity: Enables users to adjust the time range displayed.
    - **Explanation:** Added to observe historical performance and identify improvements or declines in SpaceX's launch success rates over time.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-3/6_bis_spacex_dash_app_v2.py

Predictive Analysis (Classification)

Summary of model development process:

1. **Data preparation:**
   - Selected relevant features.
   - Standardized numerical features to ensure uniformity across scales, enhancing model performance.
   - Split the dataset into training and testing sets to ensure robust model evaluation.
2. **Model building:**
   - Evaluated multiple classification algorithms:
     - Logistic Regression
     - Support Vector Machine (SVM)
     - Decision Tree
     - K-Nearest Neighbors (KNN)
   - Established a machine learning pipeline to streamline preprocessing and model training.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-4/7_SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Predictive Analysis (Classification)

3. **Model improvement:**
   - Tuned hyperparameters using **GridSearchCV** to optimize performance

4. **Model evaluation:**
   - Employed **cross-validation** techniques to assess model performance and mitigate overfitting.
   - Measured **accuracy**, **precision**, **recall**, and **F1-score** to comprehensively evaluate each model's effectiveness.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-4/7_SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Predictive Analysis (Classification)

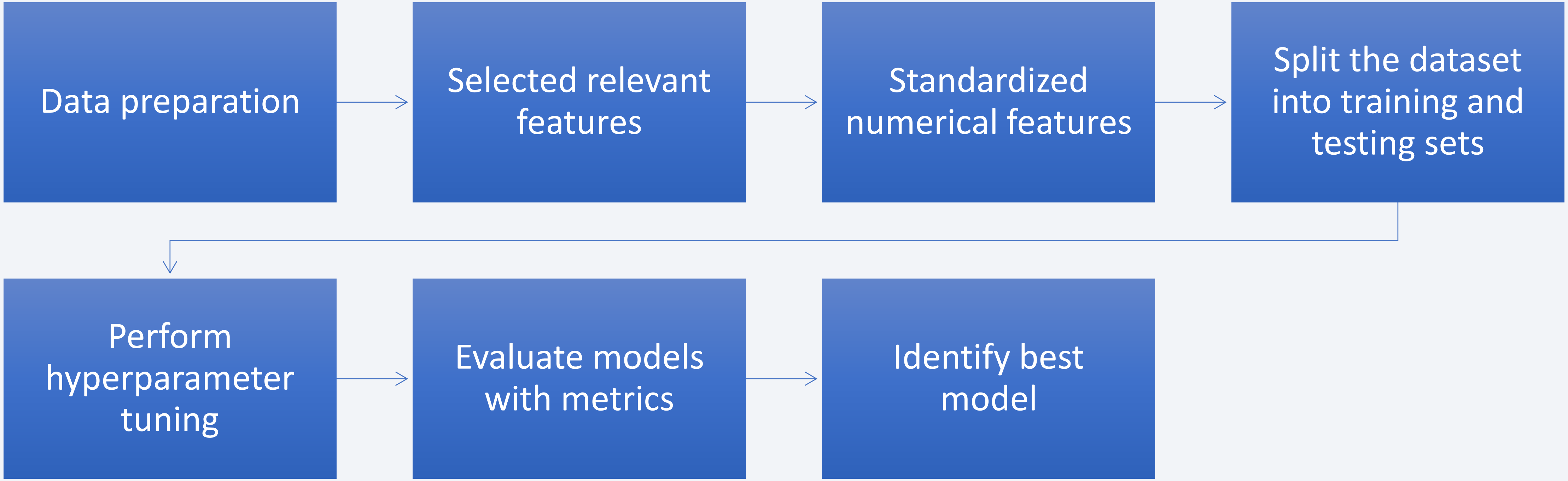
5. Best performing model:

- Among the models evaluated, the **Decision Tree classifier** was identified as the best performer with the following results:
  - Precision: 0.88 (macro average)
  - Recall: 0.88 (macro average)
  - F1-Score: 0.88 (macro average)
  - Accuracy: 0.89
- The Decision Tree model outperformed others, particularly in achieving balanced precision and recall across both classes, making it the most reliable model for predicting Falcon 9's landing success.
- Additionally, feature importance analysis revealed key variables such as payload mass and orbit type significantly influenced model predictions.

GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-4/7_SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Predictive Analysis (Classification)

Flowchart representation of model development process:

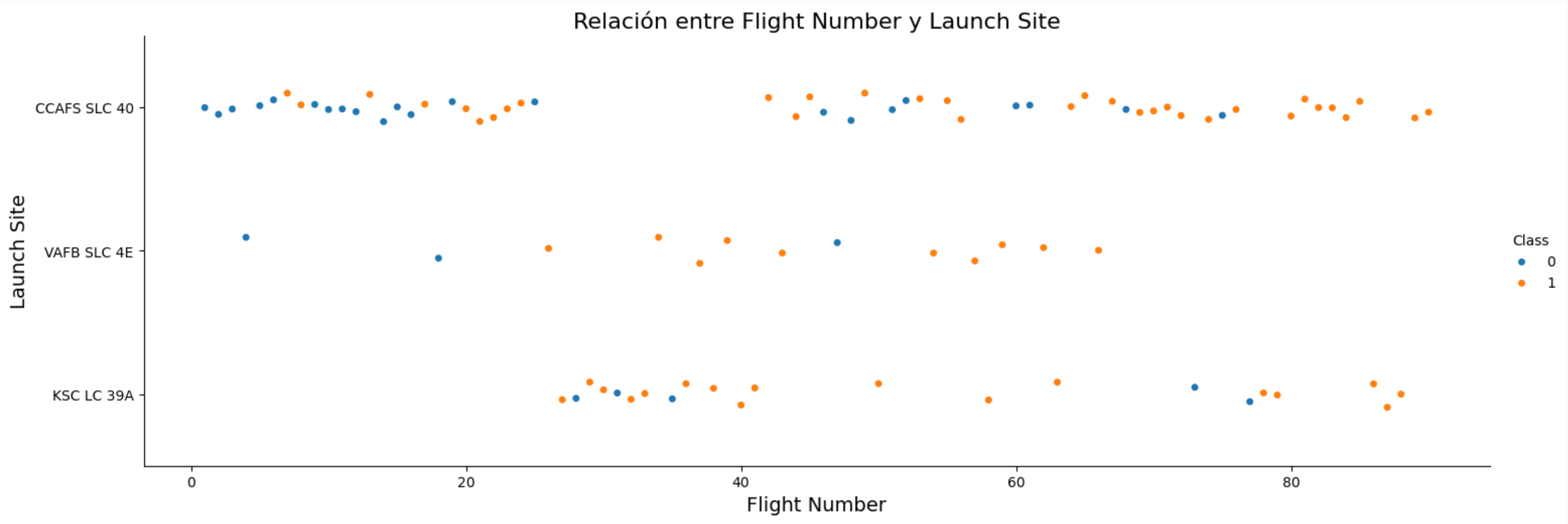


GitHub URL: https://github.com/Luar-Moon/capstone-project/blob/main/notebooks/module-4/7_SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Section 2

Insights drawn from EDA

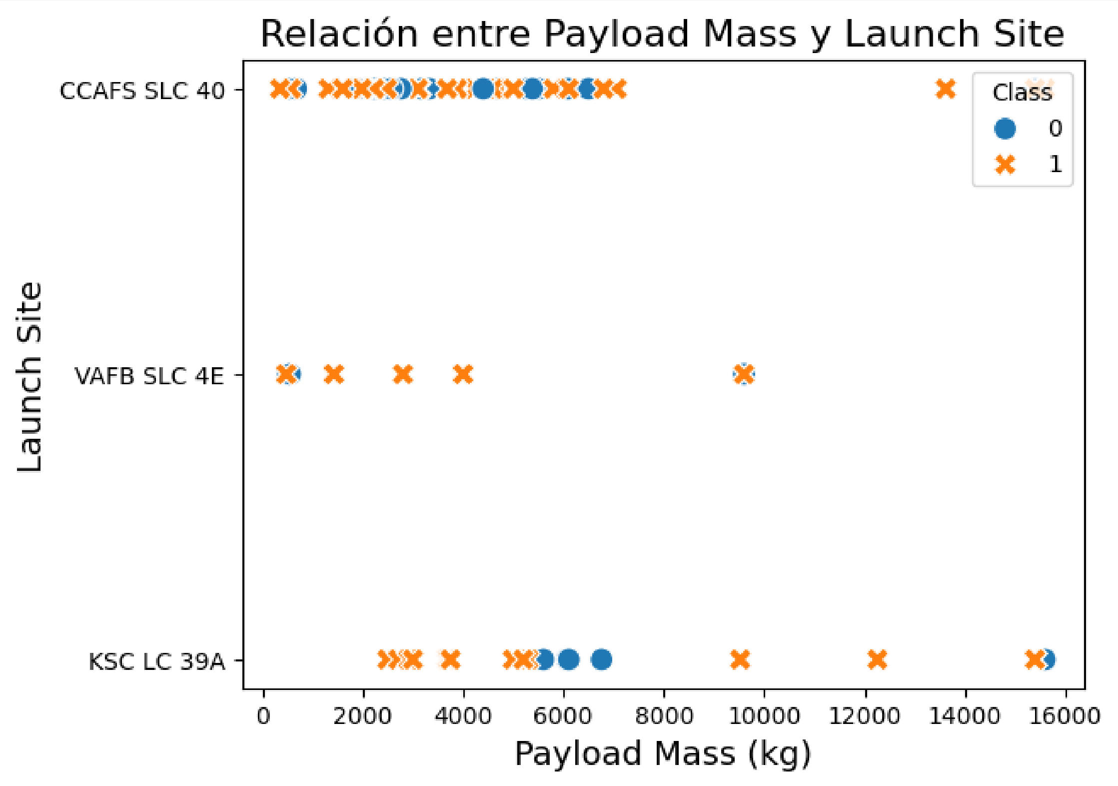
Flight Number vs. Launch Site



The scatter plot demonstrates the relationship between the sequential Flight Numbers and the corresponding Launch Sites. Each point represents a flight, with the color indicating its success or failure (Class 1 for success and Class 0 for failure).

- **General trend of improvement:** Across all sites, higher flight numbers are associated with more consistent successes, highlighting the advancements in SpaceX's launch technology and operational procedures.
- **Site-specific insights:** CCAFS SLC 40 appears to be the most utilized site with a history of consistent improvements. VAFB SLC 4E has fewer missions and a higher variability in outcomes, potentially reflecting site-specific constraints or less frequent usage. KSC LC 39A demonstrates a strong trend of recent successful launches, indicating its strategic importance in SpaceX operations.

Payload vs. Launch Site

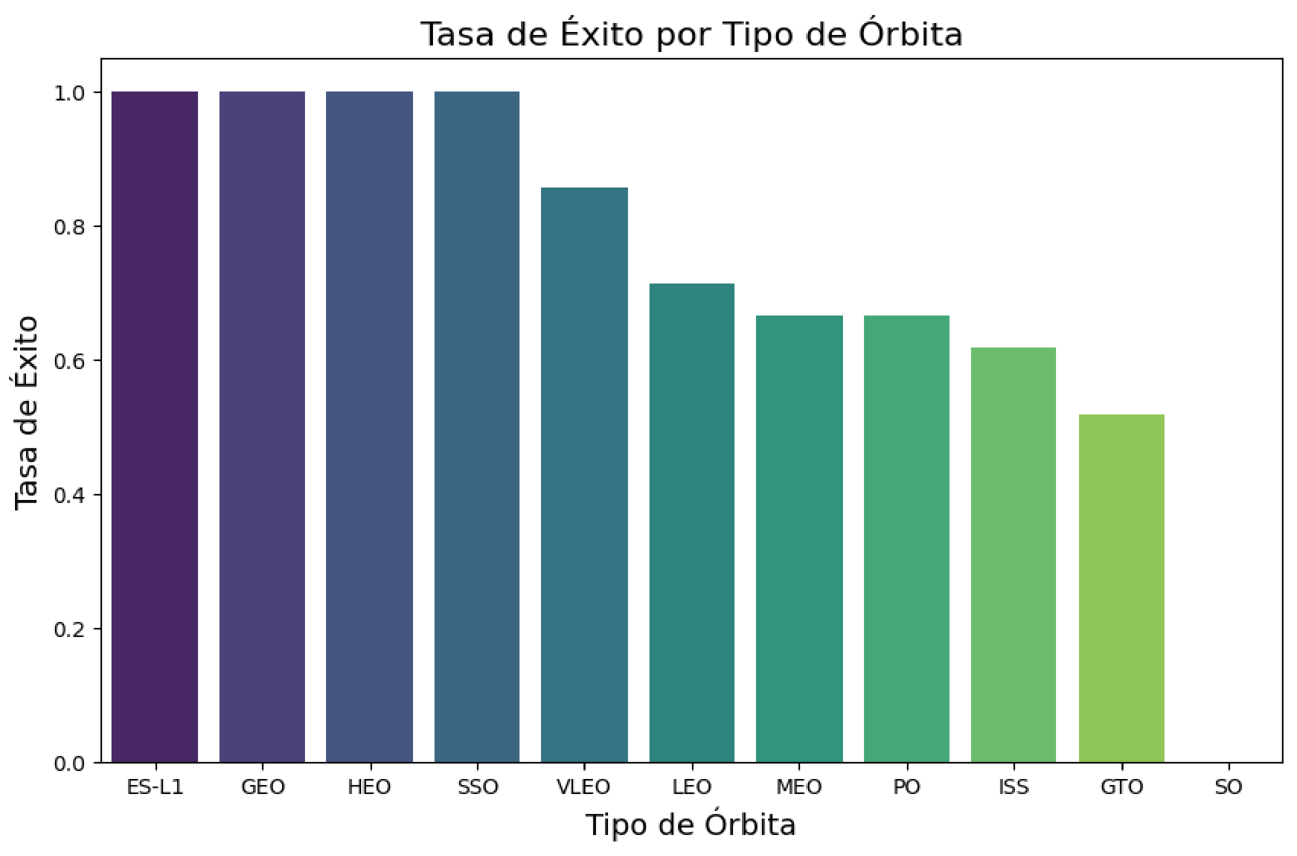


- **CCAFS SLC 40:** Handles a broad range of payloads, with most launches between 2,000 kg and 6,000 kg. Capable of heavy payloads (>10,000 kg) with a high success rate.
- **VAFB SLC 4E:** Limited to smaller payloads (<6,000 kg) with no launches above 10,000 kg, indicating operational or capacity constraints.
- **KSC LC 39A:** Supports a wide range of payloads, including heavy payloads (>10,000 kg), and shows high reliability across all payload ranges.

These insights suggest strategic differences in the utilization of launch sites, with VAFB SLC 4E focusing on smaller payloads and KSC LC 39A positioned as a hub for versatile and high-capacity launches.

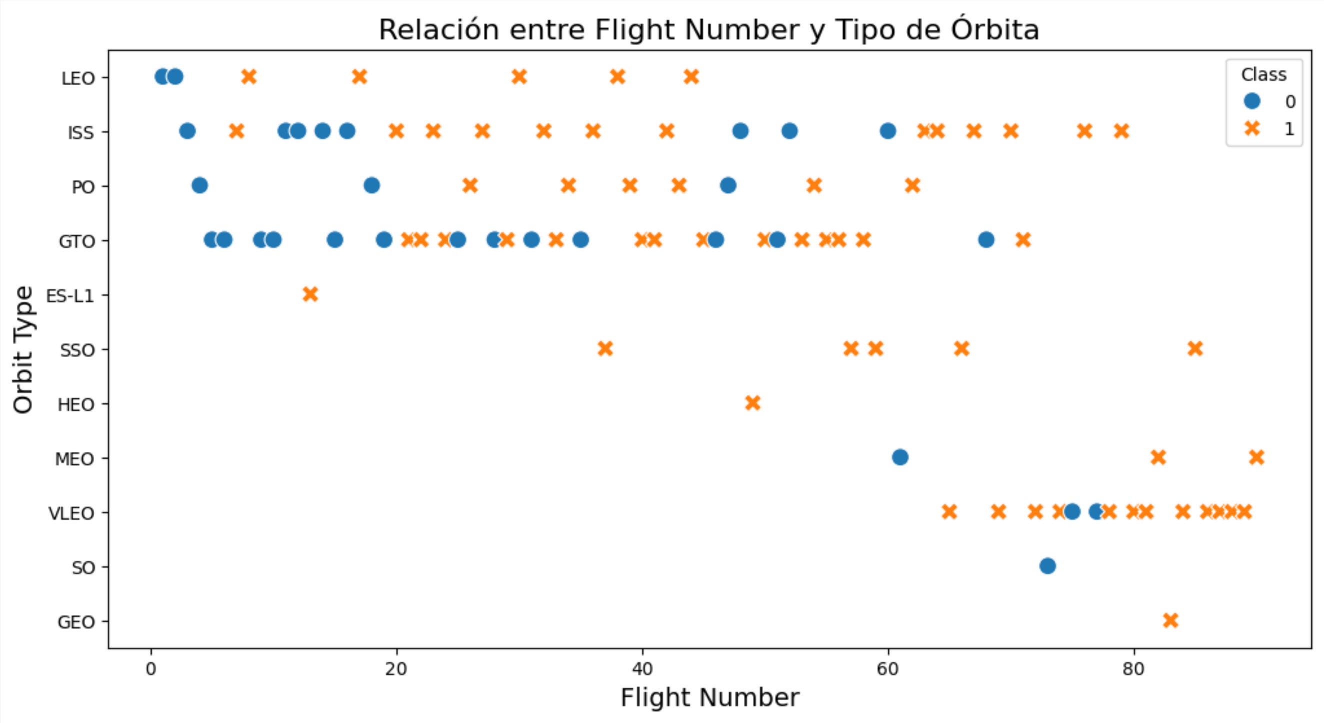
Success Rate vs. Orbit Type

- **Highest success rates:** ES-L1, GEO, and HEO orbits have a 100% success rate, reflecting high mission reliability and well-established protocols.
- **Moderate success rates:** SSO and VLEO show strong performance, commonly used for scientific and commercial purposes.
- **Lower success rates:** ISS and GTO have lower success rates, likely due to greater technical challenges. SO has the lowest success rate, indicating the complexity of these missions.



These observations emphasize SpaceX's strengths in certain orbit types while highlighting areas for improvement in more complex missions.

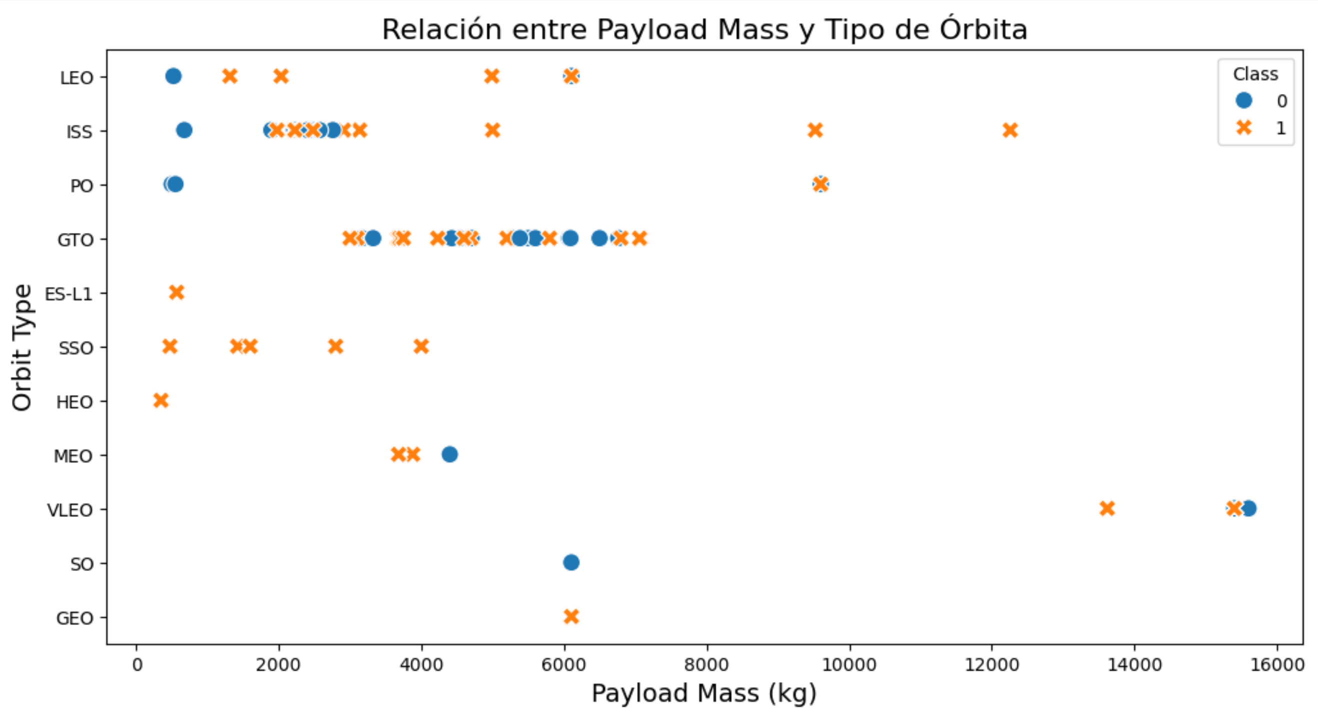
Flight Number vs. Orbit Type



- **LEO (Low Earth Orbit):** Success is strongly correlated with higher Flight Numbers, indicating operational improvements over time.
- **GTO (Geostationary Transfer Orbit):** No clear relationship between Flight Number and success, suggesting mission-specific complexities play a more significant role.
- **General insight:** Flight experience contributes to higher success rates in some orbits (e.g., LEO) but is less impactful in others (e.g., GTO).

This observations underscores the varying dynamics of success across orbit types and the role of experience and mission-specific complexities.

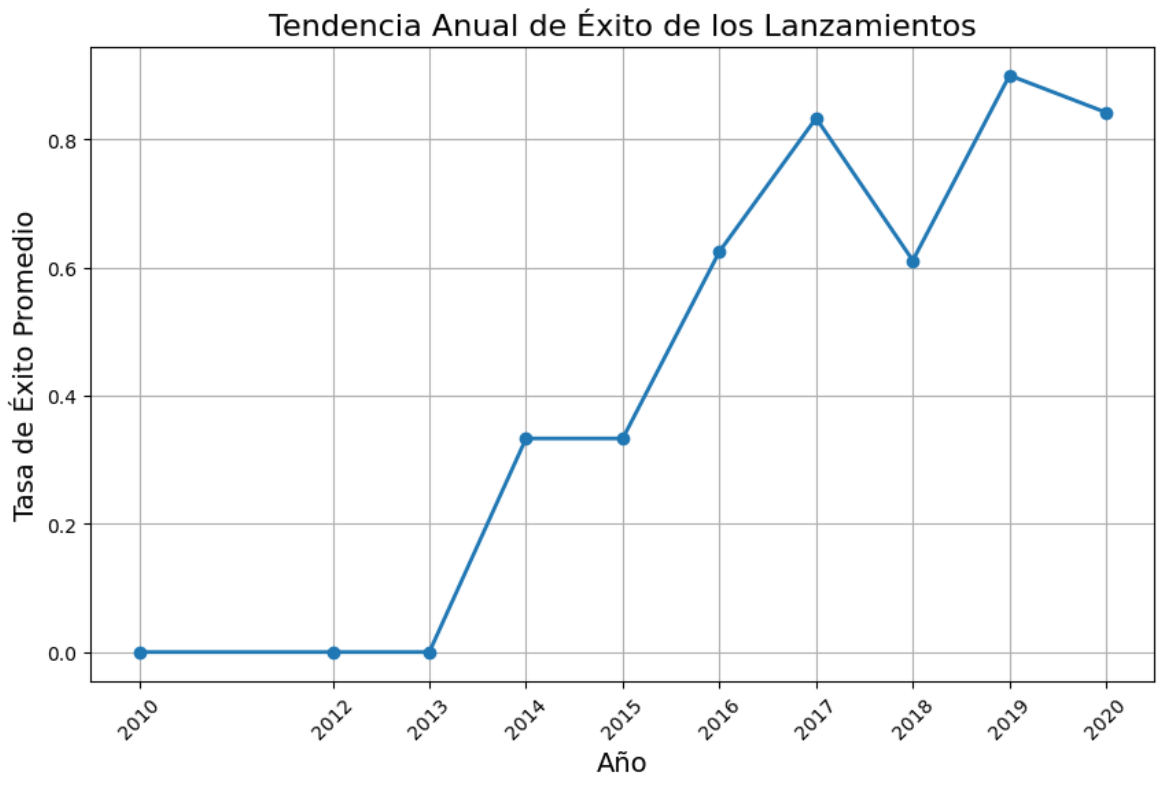
Payload vs. Orbit Type



- **Polar, LEO, and ISS orbits:** Heavy payloads (above 10,000 kg) have a higher success rate in these orbits, showcasing SpaceX's reliability for large payload missions.
- **GTO orbit:** Success and failure outcomes are mixed across payload ranges, indicating that factors beyond payload mass significantly affect mission success.
- **General trends:** Medium payloads in orbits like SSO and HEO show consistent success, while orbits like ES-L1 and GEO have limited but successful data points.

These observations underline SpaceX's strengths in specific orbits while identifying challenges in others, such as GTO.

Launch Success Yearly Trend



- **Steady growth:** SpaceX's success rate has consistently improved from 2013 to 2020, demonstrating significant operational advancements.
- **Breakthrough years:** Success rates peaked in 2017 and 2019, highlighting key milestones in reliability.
- **Early challenges:** Before 2013, success rates were at 0%, reflecting initial testing phases.

These trends showcase SpaceX's continuous progress and technological achievements over the years.

All Launch Site Names

Query:

```
result = %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
```

- The query identifies **four unique SpaceX launch sites**: CCAFS LC-40, VAFB SLC-4E, KSC LC-39A and CCAFS SLC-40.

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

- These sites are strategically located and used based on mission requirements, such as payload type and orbit destination.
- Each launch site plays a critical role in SpaceX's diverse mission portfolio, supporting both commercial and exploratory goals.

Launch Site Names Begin with 'CCA'

Query: result = %sql SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outc
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parac
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parac
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No att
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No att
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No att

- All launches were conducted at CCAFS LC-40, showcasing its importance in SpaceX's early operations.
- Missions primarily targeted Low Earth Orbit (LEO), including resupply missions to the ISS.
- Customers included NASA (COTS and CRS programs) and SpaceX for testing and qualification flights.
- Early booster versions (F9 v1.0) were used, with payloads ranging from 0 kg (test flights) to moderate masses (e.g., 677 kg).
- Mission outcomes were all successful, though landing attempts were either unsuccessful or not attempted.

This table highlights SpaceX's foundational launches and early successes at Cape Canaveral.

Total Payload Mass

Query:

```
result = %sql SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass FROM
SPACEXTBL WHERE "Customer" = 'NASA (CRS)';
```

- SpaceX delivered a **total of 45,596 kg of payload** under NASA's Commercial Resupply Services (CRS) program.

Total_Payload_Mass
45596

- These missions were vital for resupplying the International Space Station (ISS).
- The result showcases SpaceX's reliability and capability in handling critical payload deliveries for NASA.

Average Payload Mass by F9 v1.1

Query:

```
result = %sql SELECT AVG("Payload_Mass__kg_") AS Average_Payload_Mass
FROM SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1';
```

- The Falcon 9 v1.1 booster carried an **average payload of 2,928.4 kg** per mission.

Average_Payload_Mass
2928.4

- This variant marked significant advancements in payload capacity and reliability compared to earlier models.
- The result highlights its role in supporting medium-sized payloads and contributing to SpaceX's operational growth.

First Successful Ground Landing Date

Query:

```
result = %sql SELECT MIN("Date") AS First_Successful_Landing FROM SPACEXTBL
WHERE "Landing_Outcome" = 'Success (ground pad)';
```

- The first successful ground pad landing occurred on **December 22, 2015**.

First_Successful_Landing
2015-12-22

- This marked a pivotal milestone in achieving reusable rocket technology.
- The success demonstrated SpaceX's advancements in reducing launch costs and revolutionizing the aerospace industry.
- This event set a new standard for sustainable space exploration.

Successful Drone Ship Landing with Payload between 4000 and 6000

Query:

```
result = %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome"
= 'Success (drone ship)' AND "Payload_Mass__kg_" > 4000 AND "Payload_Mass__kg_" < 6000;
```

- Boosters **F9 FT B1022**, **B1026**, **B1021.2**, and **B1031.2** successfully landed on drone ships with payloads between 4,000 kg and 6,000 kg.

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- Demonstrates the Falcon 9 Full Thrust (F9 FT) family's reliability in handling medium to heavy payloads.
- Highlights SpaceX's technological advancements in successful drone ship recoveries.
- This showcases SpaceX's consistent performance with mid-range payload missions.

Total Number of Successful and Failure Mission Outcomes

Query:

```
result = %sql SELECT "Mission_Outcome", COUNT(*) AS Total_Count FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

- SpaceX achieved **98+ missions successfully completed**, demonstrating a high reliability rate.

Mission_Outcome	Total_Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- **1 in-flight failure** and **1 unclear payload status** highlight rare challenges.
- This underscores SpaceX's strong operational success with minimal failures.

Boosters Carried Maximum Payload

Query:

```
result = %sql SELECT "Booster_Version" FROM SPACEXTBL WHERE
"Payload_Mass__kg_" = (SELECT MAX("Payload_Mass__kg_") FROM SPACEXTBL);
```

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

- Multiple Falcon 9 Block 5 boosters, including **B1048.4**, **B1049.4**, and **B1051.3**, carried the maximum payload mass recorded.
- The **Falcon 9 Block 5** family showcases enhanced performance and reliability in handling heavy payloads.
- These missions highlight SpaceX's capability to meet high-demand payload requirements, reinforcing their industry leadership.
- This reflects the consistent strength and scalability of the Falcon 9 Block 5 design.

2015 Launch Records

Query:

```
result = %sql SELECT SUBSTR("Date", 6, 2) AS Month, "Booster_Version", "Launch_Site", "Landing_Outcome"
FROM SPACEXTBL WHERE "Landing_Outcome" = 'Failure (drone ship)' AND SUBSTR("Date", 1, 4) = '2015';
```

- Two drone ship landing failures occurred in **January** and **April 2015** involving boosters "F9 **v1.1 B1012**" and "**B1015**".

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- Both launches were conducted at **CCAFS LC-40**, a critical SpaceX launch site.
- These failures reflect early challenges in developing reliable drone ship landing capabilities, driving improvements in reusable rocket technology.

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Query:

```
result = %sql SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY Outcome_Count DESC;
```

Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- **"No attempt"** was the most common outcome (10 occurrences), reflecting early missions without recovery efforts.
- Successful landings included **5 drone ship** and **3 ground pad** recoveries, marking milestones in reusability.
- Failures, such as **5 drone ship** and **2 parachute failures**, highlight challenges during early development phases.
- Unique outcomes like **"Controlled (ocean)"** (3) and **"Precluded (drone ship)"** (1) indicate intentional or constrained landings.

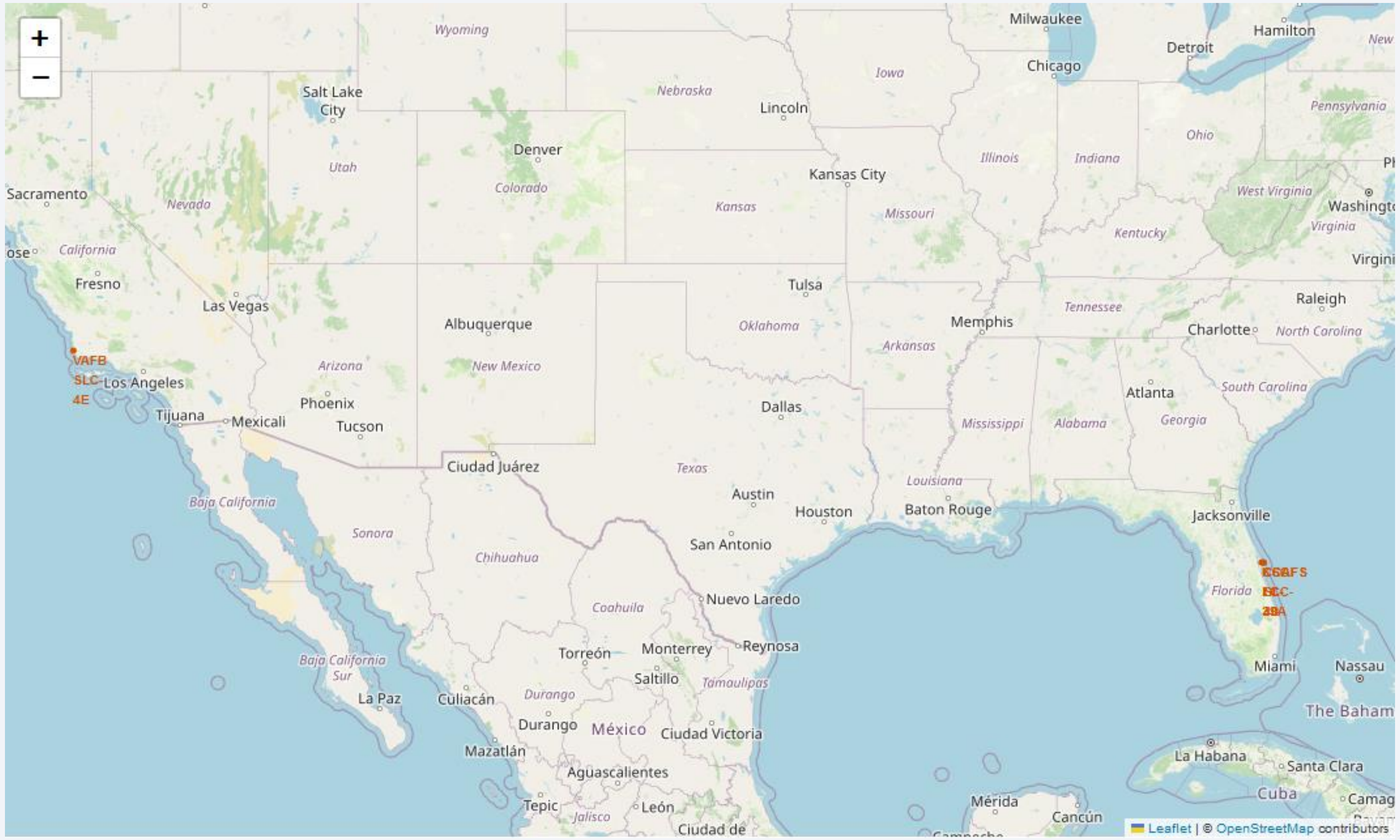
This table illustrates SpaceX’s progression toward consistent reusable rocket successes.

Section 3

Launch Sites Proximities Analysis

Geographic Locations of all Spacex Launch Sites

- The map displays all SpaceX launch sites: <u>**VAFB SLC-4E**</u> in California, and <u>**CCAFS LC-40**</u>, <u>**CCAFS SLC-40**</u>, and <u>**KSC LC-39A**</u> in Florida.
- Florida sites are strategically positioned for equatorial and geostationary orbit missions.
- California site supports polar and sun-synchronous orbits, enabling diverse mission profiles.



This distribution reflects SpaceX’s strategic planning for orbital efficiency and mission versatility.

Color-Labeled Launch Outcomes Across Sites

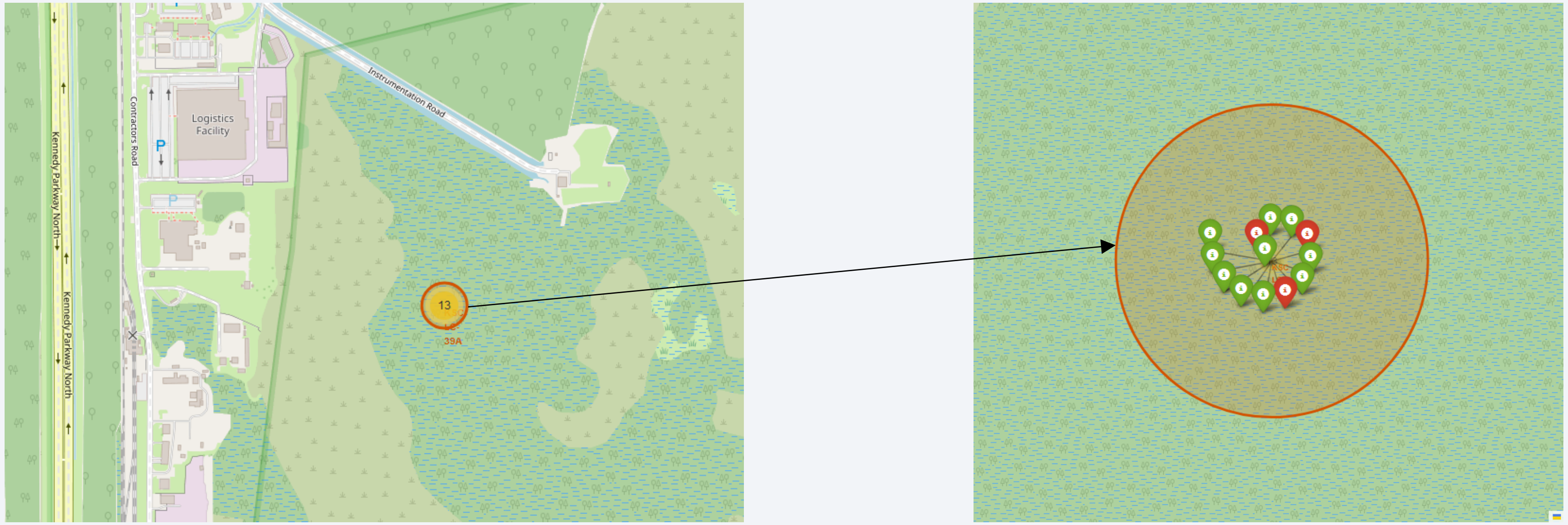
1. CCAFS SLC-40 and CCAFS LC-40

- CCAFS SLC-40 and LC-40 display a combination of green and red markers, illustrating the progression of success rates over time and emphasizing the early development phases.

[Note] Green markers indicate successful launches, while red markers show failures.

Color-Labeled Launch Outcomes Across Sites

2. <u>KSC LC-39A</u>



- KSC LC-39A exhibits a higher concentration of green markers, reflecting its strong success rate.
- Red markers are limited, showcasing the site's reliability.

[Note] Green markers indicate successful launches, while red markers show failures.

Color-Labeled Launch Outcomes Across Sites

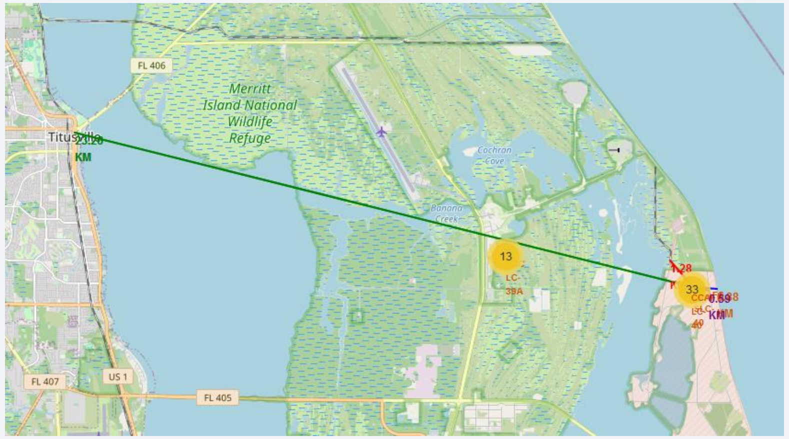
3. VAFB SLC-4E

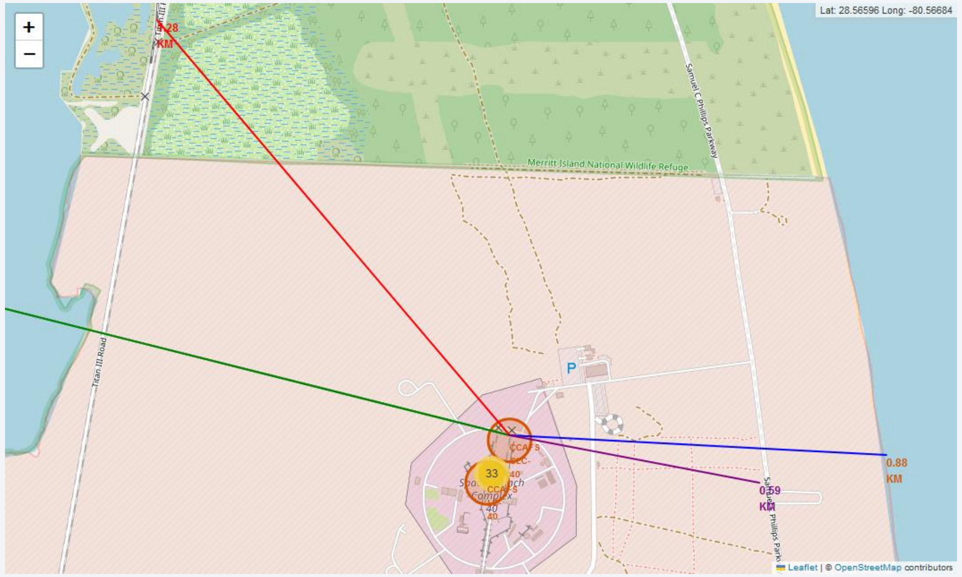
- VAFB SLC-4E markers are fewer due to its specialized use for polar orbit launches.
- Green markers dominate, highlighting the site's effectiveness for specific missions.

[Note] Green markers indicate successful launches, while red markers show failures.

Launch Site Proximities to Key Infrastructures

- **Distance from closest city – Green line (23.20 km):**
  - Provides a safety buffer for surrounding populations while minimizing the impact of launch operations on nearby cities.



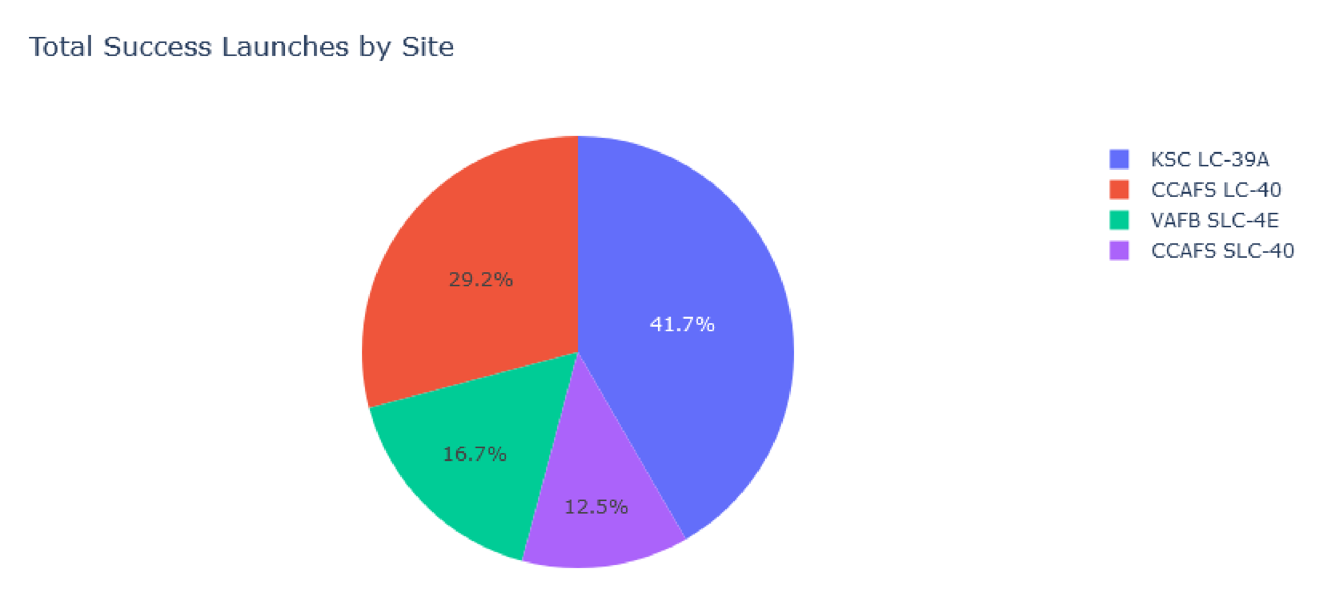


- **Coastline proximity – Blue line (0.88 km):** Ensures rockets launch safely over water, reducing risks to nearby populations and infrastructure.
- **Highway proximity – Purple line (0.59 km):** Facilitates quick and efficient transportation of resources and personnel to and from the site.
- **Railway proximity (1.28 km) – Red line:** Enables the transport of heavy equipment, such as rocket components, efficiently over long distances.

Section 4

Build a Dashboard with Plotly Dash

Proportion of Total Successful Launches by Site

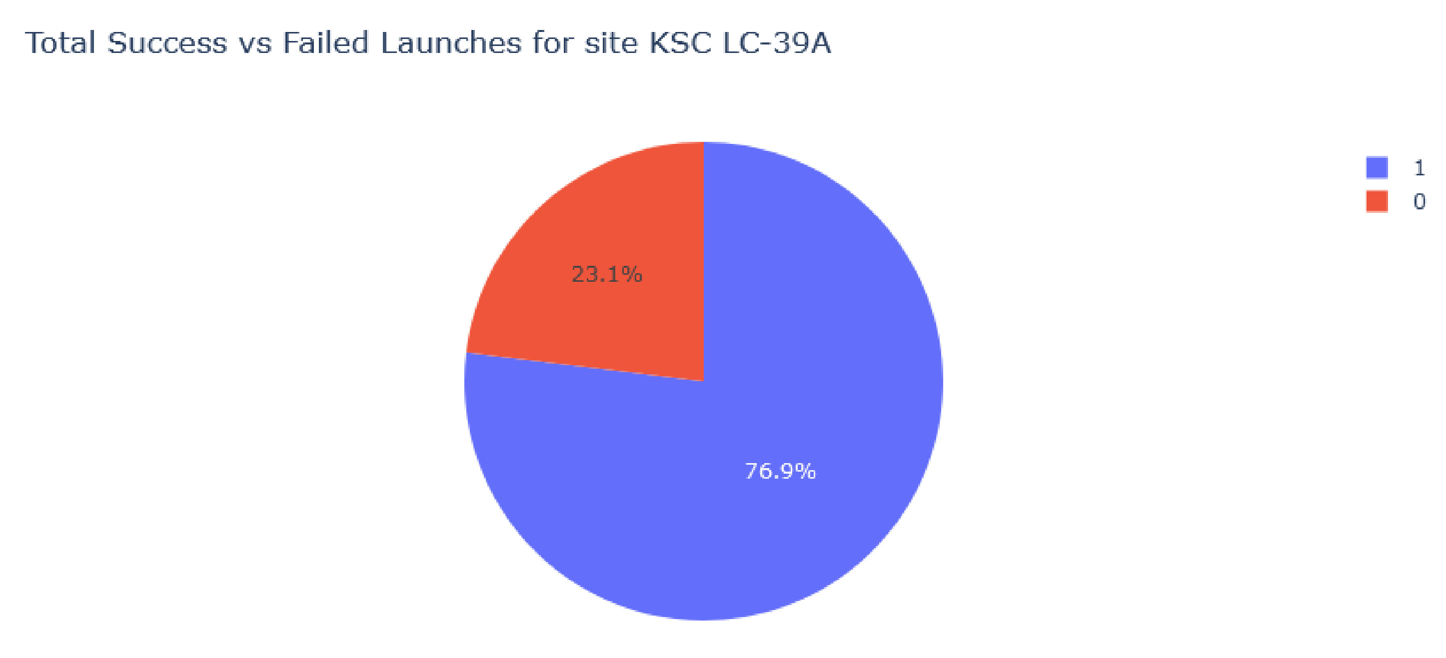


- **KSC LC-39A** accounts for the largest proportion of successful launches (41.7%), emphasizing its strategic importance.
- **CCAFS LC-40** contributes significantly with 29.2% of successful missions.
- **VAFB SLC-4E** (16.7%) and **CCAFS SLC-40** (12.5%) support a smaller but critical share of launches.

- The chart highlights the reliance on **KSC LC-39A** and **CCAFS LC-40** for most of the SpaceX's success.

This showcases the operational focus and reliability of key SpaceX launch sites.

Success vs. Failure Launch ratio for KSC LC-39A



- **KSC LC-39A** has a high success rate of **76.9%**, showcasing its reliability and efficiency.
- The **23.1% failure rate** highlights challenges faced in earlier missions or complex operations.
- This site is pivotal to SpaceX's strong record of successful launches.

The chart underscores KSC LC-39A's critical role in SpaceX's mission success.

Payload Mass vs. Launch Outcome for all sites

- **High success rate:** Success is more consistent for payloads below **6,000 kg**, across all booster versions.
- **Block 5 (B5):** Demonstrates the highest reliability, especially for heavier payloads above **6,000 kg**.
- **Older Boosters: Falcon 9 v1.1** and **FT** show lower success rates for heavier payloads but perform well for smaller payload ranges.

These observations emphasize the advancements in SpaceX's booster technology, with Block 5 leading in reliability for diverse payload requirements.

Section 5

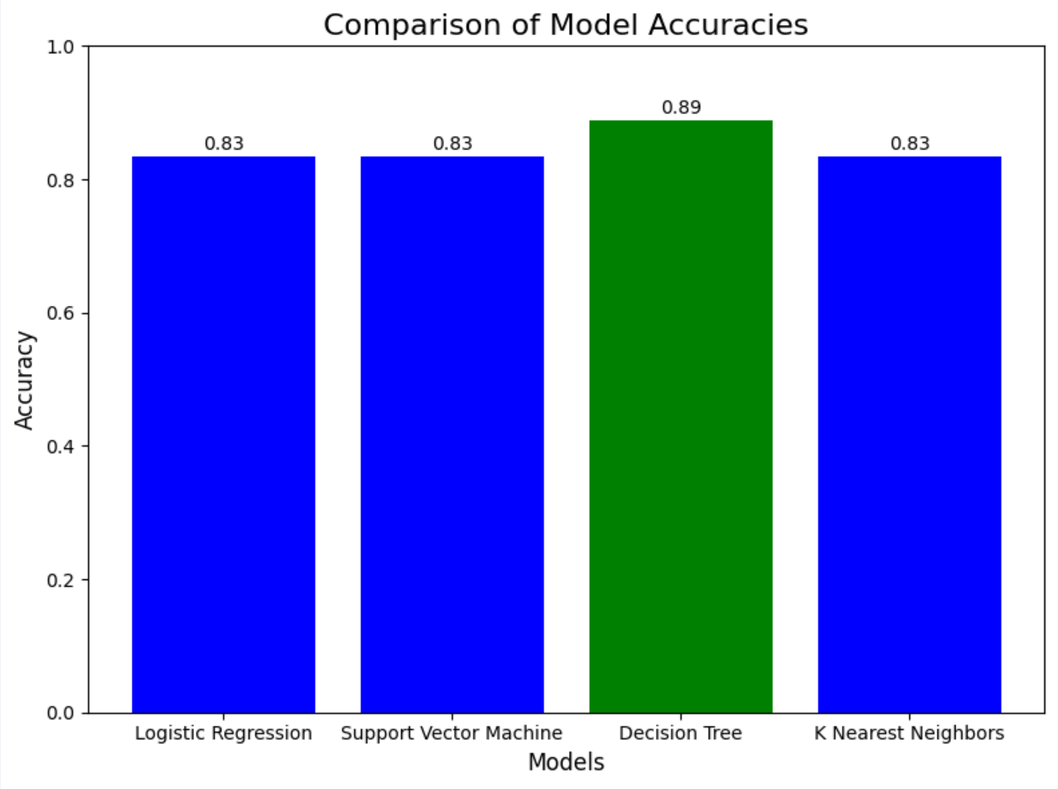
Predictive Analysis (Classification)

Classification Accuracy

Best performing model:

- Among the models compared, the Decision Tree model achieves the highest accuracy at 89%.
- Logistic Regression, Support Vector Machine, and K Nearest Neighbors share the same accuracy of 83%, making them less effective compared to the Decision Tree.

The **Decision Tree model stands out as the best-performing classifier**, providing the highest accuracy for the dataset.

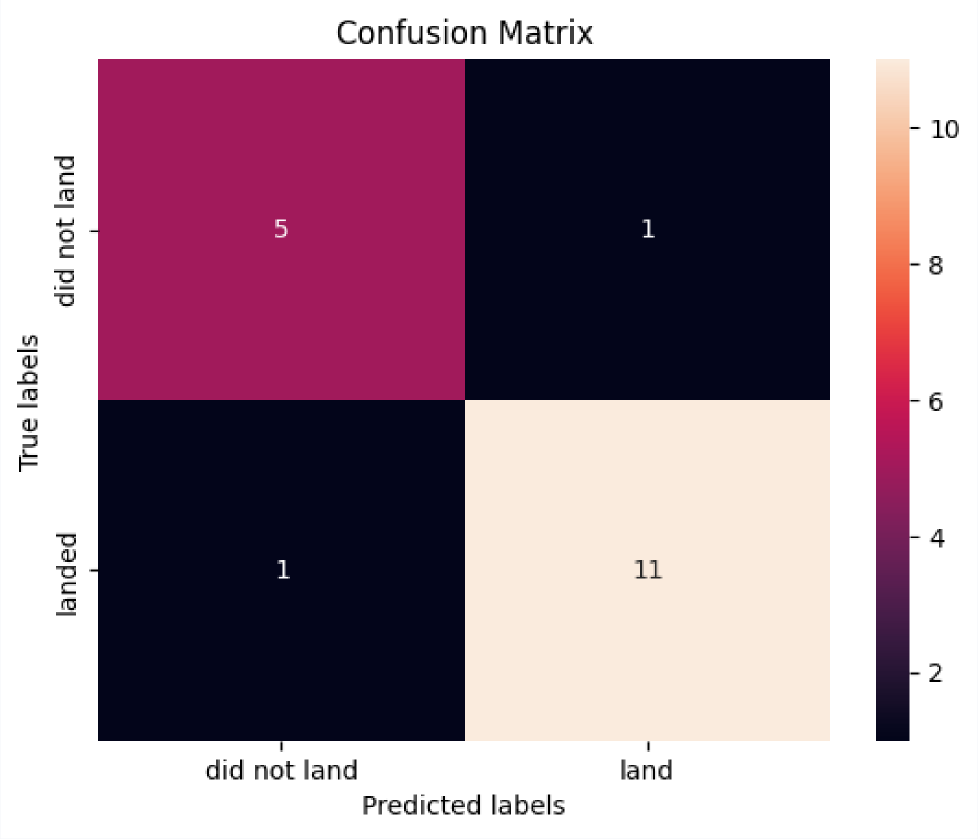


Confusion Matrix

The confusion matrix highlights the performance of the **Decision Tree model**, which achieved the highest accuracy of **89%**.

- **Correct Predictions:**
  - **11 True Positives:** Successful landings correctly predicted.
  - **5 True Negatives:** Unsuccessful landings correctly predicted.
- **Errors:**
  - **1 False Positive:** Predicted a landing success when it did not land.
  - **1 False Negative:** Missed predicting a success when the rocket actually landed.

The matrix confirms the model's ability to make accurate predictions with minimal errors (2 out of 18 predictions were incorrect), showcasing its effectiveness for the classification task.



Conclusions

- **Increasing success trends:** SpaceX has demonstrated consistent growth in success rates since 2013, achieving key milestones in 2017 and 2019, reflecting significant advancements in technology and operations.
- **Importance of launch sites:** KSC LC-39A and CCAFS LC-40 are the most reliable sites, with KSC LC-39A accounting for 41.7% of successful launches, highlighting its strategic role.
- **Payload capabilities:** Orbits like Polar, LEO, and ISS are highly reliable for heavy payloads (>10,000 kg), whereas more complex orbits such as GTO pose significant technical challenges.
- **Reusability technology:** The first successful ground pad landing in 2015 marked a milestone in cost reduction and space sustainability. Drone ship landings have shown consistent advancements with medium payloads (4,000–6,000 kg).
- **Predictive models:** The Decision Tree model was the most effective, achieving an accuracy of 89%, demonstrating a good balance of correct predictions with minimal errors.

SpaceX has showcased significant strengths in specific orbits and reusability technologies while identifying areas for improvement in more complex missions. This continuous progress underscores its leadership in the aerospace industry.

Appendix

Python code snippets

API data collection: Utilized SpaceX REST API to retrieve JSON data, parse and store it in tabular format.

- GitHub link

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Importar la librería necesaria
from pandas import json_normalize

# Decodificar el contenido JSON de la respuesta
data_json = response.json()

# Convertir la respuesta JSON en un DataFrame usando json_normalize
data = json_normalize(data_json)
```

```
# Opcional: Mostrar todas las columnas para entender los datos
data.columns
```

```
Index(['static_fire_date_utc', 'static_fire_date_unix', 'tbd', 'net', 'window',
       'rocket', 'success', 'details', 'crew', 'ships', 'capsules', 'payloads',
       'launchpad', 'auto_update', 'failures', 'flight_number', 'name',
       'date_utc', 'date_unix', 'date_local', 'date_precision', 'upcoming',
       'cores', 'id', 'fairings.reused', 'fairings.recovery_attempt',
       'fairings.recovered', 'fairings.ships', 'links.patch.small',
       'links.patch.large', 'links.reddit.campaign', 'links.reddit.launch',
       'links.reddit.media', 'links.reddit.recovery', 'links.flickr.small',
       'links.flickr.original', 'links.presskit', 'links.webcast',
       'links.youtube_id', 'links.article', 'links.wikipedia', 'fairings'],
      dtype='object')
```

Appendix

Python code snippets

Web scraping: Used Python's BeautifulSoup library to scrape additional metadata from Wikipedia.

- GitHub link

```
# URL estática de la página de Wikipedia
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

# Realiza la solicitud HTTP GET
response = requests.get(static_url)

# Verifica que la solicitud fue exitosa (código de estado 200)
if response.status_code == 200:
    print("Página descargada con éxito")
else:
    print(f"Error al descargar la página, código de estado: {response.status_code}")
```

```
Página descargada con éxito
```

Create a `BeautifulSoup` object from the HTML `response`

```
# Crea un objeto BeautifulSoup a partir del HTML
soup = BeautifulSoup(response.text, 'html.parser')
```

Appendix

Python code snippets

Machine learning pipeline: Built models like Decision Trees and performed hyperparameter tuning using GridSearchCV.

- GitHub link

```
parameters = {
    'criterion': ['gini', 'entropy'],
    'splitter': ['best', 'random'],
    'max_depth': [2 * n for n in range(1, 10)],
    'max_features': ['sqrt', 'log2', None],
    'min_samples_leaf': [1, 2, 4],
    'min_samples_split': [2, 5, 10]
}

tree = DecisionTreeClassifier()
```

```
# Crear el objeto GridSearchCV para el árbol de decisión
tree_cv = GridSearchCV(estimator=tree, param_grid=parameters, cv=10)

# Ajustar el modelo para encontrar los mejores parámetros
tree_cv.fit(X_train, Y_train)
```

```
GridSearchCV
estimator: DecisionTreeClassifier
DecisionTreeClassifier
```

```
print("tuned hpyerparameters :(best parameters) ",tree_cv.best_params_)
print("accuracy :",tree_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'criterion': 'entropy', 'max_depth': 12, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 2, 'splitter': 'best'}
accuracy : 0.8857142857142856
```

Appendix

SQL Queries

1. Calculate Total Payload for NASA CRS missions:

```
# Ejecutar la consulta SQL para calcular la masa total de carga útil en misiones NASA (CRS)
result = %sql SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass FROM SPACEXTBL WHERE "Customer" = 'NASA (CRS)';

# Mostrar el resultado
print("Masa total de carga útil en misiones NASA (CRS):")
result
```

```
 * sqlite:///my_data1.db
Done.
Masa total de carga útil en misiones NASA (CRS):
```

Total_Payload_Mass
45596

2. Identify first successful ground landing:

```
# Ejecutar la consulta SQL para obtener la primera fecha de aterrizaje exitoso en ground pad
result = %sql SELECT MIN("Date") AS First_Successful_Landing FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';

# Mostrar el resultado
print("Fecha del primer aterrizaje exitoso en ground pad:")
result
```

```
 * sqlite:///my_data1.db
Done.
Fecha del primer aterrizaje exitoso en ground pad:
```

First_Successful_Landing
2015-12-22

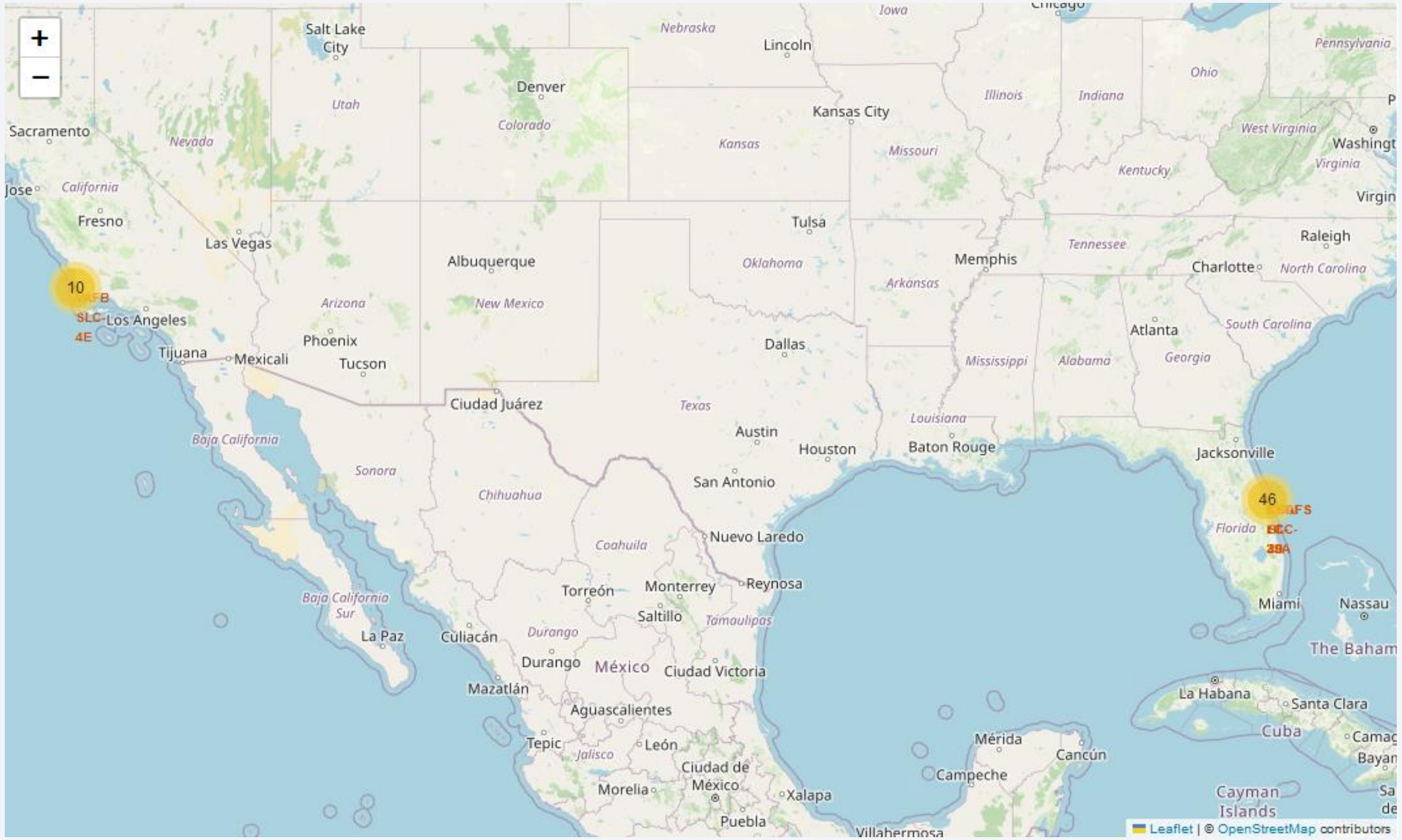
- GitHub link

Appendix

Folium Maps

Map of launch sites.

- [GitHub link]



Thank you!